(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number
WO 02/20844 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/27655

(22) International Filing Date:
7 September 2001 (07.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/231,281 8 September 2000 (08.09.2000) US

(71) Applicant: **POLYGENYX, INC.** [US/US]; Three Biotech, One Innovation Drive, Worcester, MA 01605 (US).

(72) Inventor: **LANDERS, John, E.**; 1113 Applebriar Lane, Marlboro, MA 01752 (US).

(74) Agent: **LOCKHART, Helen, C.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/20844 A1

(54) Title: METHODS AND PRODUCTS RELATED TO HIGH THROUGHPUT GENOTYPE ANALYSIS

(57) Abstract: The invention relates to high throughput methods for determining genotypes. The improved high throughput methods involve quantitative analysis of the hybridization of specific probes to each allele of a SNP.

METHODS AND PRODUCTS RELATED TO HIGH THROUGHPUT GENOTYPE ANALYSIS

BACKGROUND OF THE INVENTION

Increasing numbers of diseases have been correlated with genetic alterations in recent years. Many genetic alterations have been identified in the course of the routine screening of the genome but others have also been identified in specific genome wide scans for the identification of genetic alterations associated with diseases. In general, genome wide scans may be performed using polymorphic DNA markers to determine which markers segregate with a complex trait of interest. The loci which are identified as contributing to a disease can then be mapped to a specific genomic region based on the known chromosomal locations of the markers segregating with or "linked" to that trait. Several types of DNA polymorphisms or markers occur in the human genome and can be used in genome wide scans. These include restriction fragment length polymorphisms (RFLPs), microsatellites, or simple sequence length polymorphisms (SSLPs), and single nucleotide polymorphisms (SNPs).

SNPs are biallelic regions within a genome. That is, a SNP refers to a single base pair position in the genome at which two different sequence alternatives (alleles) exist in the population at frequencies of greater than 1%. SNPs are extremely stable and dense within the genome and are an increasingly valuable tool for analyzing genetic alternations. For instance, the low mutation rate of SNPs makes them excellent markers for studying complex genetic traits. Recently, a high throughput method for genotyping using SNPs has been identified (WO 00/18960, Landers et al.). The method described in Landers et al. relates to the use of hybridization assays for identifying SNPs in genetic samples which have been reduced in complexity to optimize the hybridization reaction.

SUMMARY OF THE INVENTION

The invention relates in some aspects to improved high throughput methods for SNP-based genotyping. Hybridization based assays have revolutionized the field of genotyping by providing high throughput systems which are able to process thousands more samples in fractions of the time required for prior art methods. Applicants have

surprisingly discovered a method for improving on these high throughput hybridization assays to increase the accuracy of the analysis as well as the throughput time and processing steps.

Thus, in one aspect, the invention relates to a method for analyzing a genotype by contacting a DNA fragment with two allele specific oligonucleotide (ASO) probes, each ASO probe representing one allele of a SNP, and each ASO probe having a different label, and measuring the ratio of the ASO probes specifically hybridized to the DNA fragment to analyze the genotype. In one embodiment, an approximately equivalent amount of the two ASO probes are used.

The method may be performed in solution or optionally the DNA fragment may be fixed to a surface, such as, for example, a multiwell dish, a chip, a slide, or a bead.

The two ASO probes may be hybridized with the DNA fragment at the same or separately. Thus, in some embodiments, each of the ASO probes is hybridized to the DNA fragment simultaneously with one another. In other embodiments, the ASO probes are hybridized to the DNA fragment consecutively.

The ASO probes used in the hybridization method are labelled. In one embodiment, at least one of the ASO probes contains a fluorescent label or quencher. In other embodiments, both of the ASO probes contain a fluorescent label or quencher, the fluorescent label or quencher of the two ASO probes, being distinct from one another.

The DNA fragment used in the method is prepared by any method known in the art. In some embodiments, the DNA fragment is prepared by PCR amplification of a polymorphic locus from a genomic DNA sample. In other embodiments, the DNA fragment is a reduced complexity genome. In yet other embodiments, the DNA fragment is genomic DNA or mitochondrial genomic DNA.

In other aspects, the invention relates to a method for analyzing the genotype by performing two hybridization reactions. The method is performed by contacting a first DNA fragment with a first ASO probe representing one allele of a SNP, contacting a second DNA fragment with a second ASO probe representing a second allele of the SNP, and measuring the ratio of the ASO probes specifically hybridized to the DNA fragment to analyze the genotype. In one embodiment, an approximately equivalent amount of the

two ASO probes is used in the hybridization. The first and second DNA fragments both contain the same SNP.

The hybridization reactions of the method can be performed in solution or on a surface. Thus, in some embodiments, the first and second DNA fragments are fixed to a surface, such as a multiwell dish, a chip, a slide, or a bead. In one embodiment, the reaction is performed on a chip and the first and second DNA fragments are fixed on discrete locations of the chip and each of the ASO probes is hybridized to the DNA fragments simultaneously with one another. In another embodiment, the first and second DNA fragments are fixed on discrete locations of the chip and the ASO probes are hybridized to the DNA fragments consecutively.

In another aspect the invention is a method for analyzing a genotype. The method involves the steps of contacting a DNA fragment with at least two ASO probes, each ASO probe capable of specifically hybridizing to a distinct SNP, and each ASO probe having a spectrally distinct label, and detecting the presence or absence of the hybridized label to analyze the genotype.

Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combination of elements, can be included in each aspect of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a digital image of a microarray hybridized with a single ASO probe specific for an allele of a SNP. The ASO probe is labeled with a fluorescent marker that is cable of emitting light in the red wavelength.

Figure 2 is a digital image of a duplicate microarray hybridized with a single ASO probe specific for the other allele of the SNP being analyzed in Figure 1. The ASO probe is labeled with a fluorescent marker that is cable of emitting light in the green wavelength.

Figure 3 is a computer overlay image of the images of the microarrays from Figure 1 and Figure 2.

Figure 4 is a computer generated image depicting the ratio of fluorescent label detected from the first and second ASO probes, following the overlay shown in Figure 3. The regions of the image consisting of entirely green circles correspond to a DNA sample which is homozygous for one allele. Those regions of the image consisting of entirely red circles correspond to a DNA sample which is homozygous for the other allele. The mixed color circles correspond to a DNA sample which is heterozygous.

Figure 5 is a histogram depicting the quantitative values of the ratio of the fluorescent labels from the two ASO probes. The values ranging from 10^{-1} to 10^{1} represent the DNA fractions which are heterozygous. The values ranging from 10^{-4} to $10^{-1.5}$ represent the DNA fractions which are homozygous for one allele. The values ranging from 10^{1} to 10^{3} represent the DNA fractions which are homozygous for the other allele.

DETAILED DESCRIPTION

The invention is based in part on the surprising discovery that the accuracy of SNP-based high throughput screening assays can be improved by identifying a quantitative measurement indicative of the presence or absence of each allele of a SNP. High throughput methods depend, generally, on the ability to analyze large numbers of data rapidly and efficiently. The ability to control the accuracy of the readings in a high throughput system is essential. The methods of the invention provide improved systems for controlling the accuracy of high throughput SNP screening methods for analyzing genotypes.

In one aspect, the invention is based on the discovery that a quantitative determination of the results of an allele specific oligonucleotide (ASO) hybridization reaction provides a more accurate indication of the genotype in these high throughput systems. The method involves performing a hybridization reaction between a DNA fragment and two ASO probes, each ASO probe representing one allele of a SNP. The ratio of the ASO probes which specifically hybridize to the DNA fragment is then measured. The ratio of the probes which have hybridized provides a more accurate representation of the identity of the allele within the DNA fragment than a simple qualitative determination of whether an ASO representative of an allele has hybridized or not, i.e., a binary determination.

The methods are useful for analyzing genotypes. "Genotyping" is the process of identifying the presence or absence of specific genomic sequences within genomic DNA. Distinct genomes may be isolated from individuals of populations which are related by some phenotypic characteristic, by familial origin, by physical proximity, by race, by class, etc. in order to identify polymorphisms (e.g. ones associated with a plurality of distinct genomes) which are correlated with the phenotype family, location, race, class, etc. Alternatively, distinct genomes may be isolated at random from populations such that they have no relation to one another other than their origin in the population. Identification of polymorphisms in such genomes indicates the presence or absence of the polymorphisms in the population as a whole, but not necessarily correlated with a particular phenotype.

Although genotyping is often used to identify a polymorphism associated with a particular phenotypic trait, this correlation is not necessary. Genotyping only requires that a polymorphism, which may or may not reside in a coding region, is present. When genotyping is used to identify a phenotypic characteristic, it is presumed that the polymorphism affects the phenotypic trait being characterized. A phenotype may be desirable, detrimental, or, in some cases, neutral.

Polymorphisms identified according to the methods of the invention can contribute to a phenotype. Some polymorphisms occur within a protein coding sequence and thus can affect the protein structure, thereby causing or contributing to an observed phenotype. Other polymorphisms occur outside of the protein coding sequence but affect the expression of the gene. Still other polymorphisms merely occur near genes of interest and are useful as markers of that gene. A single polymorphism can cause or contribute to more than one phenotypic characteristic and, likewise, a single phenotypic characteristic may be due to more than one polymorphism. In general multiple polymorphisms occurring within a gene correlate with the same phenotype. Additionally, whether an individual is heterozygous or homozygous for a particular polymorphism can affect the presence or absence of a particular phenotypic trait.

Phenotypic correlation is performed by identifying an experimental population of subjects exhibiting a phenotypic characteristic and a control population which do not exhibit that phenotypic characteristic. Polymorphisms which occur within the

experimental population of subjects sharing a phenotypic characteristic and which do not occur in the control population are said to be polymorphisms which are correlated with a phenotypic trait. Once a polymorphism has been identified as being correlated with a phenotypic trait, genomes of subjects which have potential to develop a phenotypic trait or characteristic can be screened to determine occurrence or non-occurrence of the polymorphism in the subjects' genomes in order to establish whether those subjects are likely to eventually develop the phenotypic characteristic. These types of analyses are generally carried out on subjects at risk of developing a particular disorder such as Huntington's disease or breast cancer.

A phenotypic trait encompasses any type of genetic disease, condition, or characteristic, the presence or absence of which can be positively determined in a subject. Phenotypic traits that are genetic diseases or conditions include multifactorial diseases of which a component may be genetic (e.g. owing to occurrence in the subject of a SNP), and predisposition to such diseases. These diseases include, but not limited to, asthma, cancer, autoimmune diseases, inflammation, blindness, ulcers, heart or cardiovascular diseases, nervous system disorders, and susceptibility to infection by pathogenic microorganisms or viruses. Autoimmune diseases include, but are not limited to, rheumatoid arthritis, multiple sclerosis, diabetes, systemic lupus, erythematosus and Grave's disease. Cancers include, but are not limited to, cancers of the bladder, brain, breast, colon, esophagus, kidney, hematopoietic system e.g. leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach, and uterus. A phenotypic characteristic includes any attribute of a subject other than a disease or disorder, the presence or absence of which can be detected. Such characteristics can, in some instances, be associated with occurrence of a SNP in a subject which exhibits the characteristic. Examples of characteristics include, but are not limited to, susceptibility to drug or other therapeutic treatments, appearance, height, color (e.g. of flowering plants), strength, speed (e.g. of race horses), hair color, etc. Many examples of phenotypic traits associated with genetic variation have been described, see e.g., US Patent No. 5,908,978 (which identifies association of disease resistance in certain species of plants associated with genetic variations) and US Patent No. 5,942,392 (which describes genetic markers associated with development of Alzheimer's disease).

Identification of associations between genetic variations (e.g. occurrence of SNPs) and phenotypic traits is useful for many purposes. For example, identification of a correlation between the presence of a SNP allele in a subject and the ultimate development by the subject of a disease is particularly useful for administering early treatments, or instituting lifestyle changes (e.g., reducing cholesterol or fatty foods in order to avoid cardiovascular disease in subjects having a greater-than-normal predisposition to such disease), or closely monitoring a patient for development of cancer or other disease. It may also be useful in prenatal screening to identify whether a fetus is afflicted with or is predisposed to develop a serious disease. Additionally, this type of information is useful for screening animals or plants bred for the purpose of enhancing or exhibiting of desired characteristics.

The genotype is analyzed using standard hybridization methods. In general, a DNA fragment is contacted with two ASO probes, each one representing one allele of a SNP. The method may be performed by contacting a single DNA fragment, in solution or on a surface, with both ASO probes, each one having a different label to distinguish it. Alternatively, the DNA fragment may be separated into two samples and each sample hybridized with one of the probes. If the two sample method is used, the two ASO probes may have the same labels or different labels. The amount of ASO probe which has hybridized to the DNA fragment can be measured and quantitated. The ratio of the two amounts can then be determined by comparing the quantitative value for each probe.

The ratio of the quantitative value may be used to identify an accurate determination of the genotype. An example of the ratios indicating whether an individual is homozygous for one or the other allele or is heterozygous is shown in the histogram of Figure 5. The histogram of Figure 5 demonstrates three discrete peaks. The central peak is representative of DNA samples containing both alleles of the SNP. The peak to the left is indicative of a DNA sample containing one allele and the peak to the right is indicative of a DNA sample containing the other allele. Thus if a sample has a ratio falling within the region of the central peak the sample is indicative of a heterozygous individual. If the sample has a ratio falling within either the left or the right peak the sample is indicative of a homozygous individual.

The actual values depicted in the population histogram of Figure 5, are representative and serve as accurate indicators of experiments carried out under similar experimental conditions. Under these types of conditions, the values ranging from 10^{-1} to 10^{1} represent the DNA fractions which are heterozygous. The values ranging from 10^{-4} to 10^{-1} represent the DNA fractions which are homozygous for one allele. The values ranging from 10^{1} to 10^{3} represent the DNA fractions which are homozygous for the other allele.

The actual quantitative value of the ratios obtained according to the methods of the invention will vary somewhat depending on experimental factors, such as the type of computer program used and the type of fluorescent marker. These types of factors can be adequately controlled by performing a control analysis under the particular experimental conditions using a known set of research tools. For instance, some computer programs provide better discrimination between background noise and actual signal, i.e. by removing from the analysis pixels that have extremely high or low values. The better the computer program at providing higher discrimination, the higher the ratio produced and the greater the shift of the data in either direction (i.e. more discrimination between the three peaks). Computer programs having good discrimination include but are not limited to Imagene (Biodiscovery), Arrayvision, and Autogene (Biodiscovery). Computer programs which are less able to distinguish noise from signal have less shift and thus the three peaks obtained are closer to one another and the ratios are smaller. These types of computer programs are still useful according to the methods of the invention, but the difference in absolute values of the ratios should be taken into consideration. Examples of these programs include but are not limited to Scanalyze and Arrayworx (Applied Precision).

The type of fluorophores used will also have some influence on the absolute values of the ratios. In general bright, highly stable fluorophores such as Cy3 will produce higher ratios similar to highly discriminating computer programs. Fluorophores such as Cy5 which are less bright and more susceptible to degradation tend to produce lower ratios. These properties of fluorophores are well known to those of skill in the art and are often characterized in the product literature accompanying a fluorophore when it is purchased from a commercial source.

The methods of the invention involve analysis of both alleles of a SNP. A "single nucleotide polymorphism" or "SNP" as used herein is a single base pair (i.e., a pair of complementary nucleotide residues on opposite genomic strands) with a DNA region wherein the identities of the paired nucleotide residues vary from individual-to-individual. At the variable base pair (alleles) in the SNP two or more alternative base pairings can occur at a relatively high frequency in a subject (e.g., human) population. Many methods have been described for identifying SNPs. (see e.g. W095/12607, Bostein, et al., *Am. J. Hum. Genet,*.32:314-331 (1980), etc.).

A "polymorphic region" is a region or segment or DNA, the nucleotide sequence of which varies from individual-to-individual. The two DNA strands which are complementary to one another except at the variable positions are referred to alleles. A polymorphism is allelic because some members of a species have one allele and other members have a variant allele and some have both. When only one variant sequence exists, a polymorphism is referred to as a diallelic polymorphism. There are three possible genotypes in a diallelic polymorphic DNA in a diploid organism. These three genotypes arise because it is possible that a diploid individuals DNA may be homozygous for one allele, homozygous for the other allele, or heterozygous (i.e., having one copy of each allele). When other mutations are present, it is possible to have triallelic or higher order polymorphisms. These multiple mutation polymorphisms produce more complicated genotypes.

SNPs are well-suited for studying sequence variation because they are relatively stable (i.e. they exhibit low mutation rates) and because it appears that SNPs can be responsible for inherited traits. These properties make SNPs particularly useful as genetic markers for identifying disease-associated genes.

The SNPs are analyzed using an ASO hybridization reaction in which two ASOs (corresponding to each allele of a SNP are hybridized with a DNA fragment. Either the ASOs or the DNA fragment can, optionally be immobilized on a surface. The genotype is determined based on a quantitative analysis of the ratio of ASO hybridization with at least some of the DNA fragment.

In some embodiments, the hybridization step is carried out on a surface. In these embodiments, either the DNA fragment or the ASO probe can be fixed to the surface.

Preferably, automated hybridization for genotyping analysis is performed. In general, the DNA fragment is generated, e.g., isolation of genomic DNA, preparation of reduced complexity genomes, or preparation of PCR amplified DNA sequences. The DNA fragment may be spotted onto one or more surfaces (e.g., multiple glass slides) for genotyping. This process can be performed using a microarray spotting apparatus which can spot more than 1,000 samples within a square centimeter area, or more than 10,000 samples on a typical microscope slide. Each slide may be hybridized with a fluorescently tagged ASO under standard TMAC conditions analogous to those described in co-pending U.S. Patent Application No. 09/404,912, filed on September 24, 1999 (and related PCT publication WO 00/18960). The genotype of each individual can be determined by detecting the intensity of the probe for each ASO and identifying the ratio of the two.

The DNA fragment may be attached to the slide by any method for attaching DNA to a surface. Methods for immobilizing nucleic acids have been described extensively, e.g., in US Patent Nos. 5,679,524; 5,610,287; 5,919,626; and 5,445,934. For instance, DNA fragments may be spotted onto poly-L-lysine- coated glass slides, and then crosslinked by UV irradiation. A second, more preferred method, which has been developed, involves including a 5′ amino group on each of the DNA fragments. The DNA fragments are spotted onto silane-coated slides in the presence of NaOH in order to covalently attach the fragments to the slide. This method is advantageous because a covalent bond is formed between the fragments and the surface. Another method for accomplishing DNA fragment immobilization is to spot the DNA fragments onto a nylon membrane. Other methods of binding DNA to surfaces are possible and are well known to those of ordinary skill in the art. For instance, attachment to amino-alkyl-coated slides can be used. More detailed methods are described in co-pending U.S. Patent Application No. 09/404,912, filed on September 24, 1999 (and related PCT publication WO 00/18960).

The surface to which the oligonucleotide arrays are conjugated is preferably a rigid or semi-rigid support which may, optionally, have appropriate light absorbing or transmitting characteristics for use with commercially available detection equipment. Substrates which are commonly used and which have appropriate light absorbing or

transmitting characteristics include, but are not limited to, glass, Si, Ge, GaAs, GaP, SiO_2, SiN_4, modified silicon, and polymers such as (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof. Additionally, the surface of the support may be non-coated or coated with a variety of materials. Coatings include, but are not limited to, polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, and membranes. A surface, as used herein, refers to any type of solid support material to which a molecular component such as an ASO probe or DNA fragment is capable of being fixed. Surfaces include, for instance, single or multiwell dishes, chips, slides, membranes, beads, agarose, or other types of solid support mediums. Methods for fixing DNA on a surface are well known to those of ordinary skill in the art and have been described in publications such as, for instance, co-pending U.S. Patent Application No. 09/404,912, filed on September 24, 1999 (and related PCT publication WO 00/18960).

Thus, in one embodiment. the ASOs are hybridized under standard hybridization conditions with DNA fragments covalently conjugated to a surface. Briefly, ASOs are labeled at their 5′ ends. A hybridization mixture containing the ASOs and, optionally, an isostabilizing agent, denaturing agent, or renaturation accelerant is brought into contact with an array of DNA fragments immobilized on the surface and the mixture and the surface are incubated under appropriate hybridization conditions. The ASOs which do not hybridize are removed by washing the array with a wash mixture (such as a hybridization buffer) to leave only hybridized ASOs attached to the surface. After washing, detection of the label (e.g., a fluorescent molecule) is performed. For example, an image of the surface can be captured (e.g., using a fluorescence microscope equipped with a CCD camera and automated stage capabilities, phosphoimager, etc.). The label may also, or instead, be detected using a microarray scanner (e.g. one made by Genetic Microsystems). A microarray scanner provides image analysis which can be converted to a binary (i.e. +/-) signal for each sample using, for example, any of several available software applications (e.g., NIH image, ScanAnalyze, etc.) in a data format. The intensity of the signal is then measured.

Conditions for optimal hybridization will vary depending on the particular reaction, but such parameters are know to those of skill in the art. In general, the ASO is

present in a hybridization mixture at a concentration of from about 0.005 nanomoles per liter ASO hybridization mixture to about 50 nM ASO per ml hybridization mixture. More preferably, the concentration is from 0.5 nanomoles per liter to 1 nanomole per liter. A preferred concentration for radioactivity is 0.66 nanomoles per liter. The mixture preferably also includes a hybridization optimizing agent in order to improve signal discrimination between genomic sequences which are identically complementary to the ASO and those which contain a single mismatched nucleotide (as well as any neutral base etc. substitutions). Isostabilizing agents are compounds such as betaines and lower tetraalkyl ammonium salts which reduce the sequence dependence of DNA thermal melting transitions. These types of compounds also increase discrimination between matched and mismatched ASOs/DNA fragments. A denaturing agent may also be included in the hybridization mixture. A denaturing agent is a composition that lowers the melting temperature of double stranded nucleic acid molecules, generally by reducing hydrogen bonding between bases or preventing hydration of nucleic acid molecules. Denaturing agents are well-known in the art and include, for example, DMSO, formaldehyde, glycerol, urea, formamide, and chaotropic salts. The hybridization conditions in general are those used commonly in the art, such as those described in Sambrook et al., "Molecular Cloning: A Laboratory Manual", (1989), 2nd Ed., Cold Spring Harbor, NY; Berger and Kimmel, "Guide to Molecular Cloning Techniques", *Methods in Enzymology*, (1987), Volume 152, Academic Press, Inc., San Diego, CA; and Young and Davis, (1983), *PNAS* (USA) 80:1194.

In general, incubation temperatures for hybridization of nucleic acids range from about 20°C to 75°C. For probes 17 nucleotide residues and longer, a preferred temperature range for hybridization is from about 50°C to 54°C. The hybridization temperature for longer probes is preferably from about 55°C to 65°C and for shorter probes is less than 52°C. Rehybridization may be performed in a variety of time frames. Preferably, hybridization of ASO and DNA fragments is performed for at least 30 minutes.

An "allele-specific oligonucleotide" or "ASO" is an oligonucleotide which is complementary to a sequence spanning a SNP of a polymorphic region of DNA. The ASO probes used in the hybridization reaction are generally identical to one another

except for one nucleotide which is different. The nucleotide which is different in each of the ASO probes is one which is complementary to each of the two alleles of the SNP. Thus, for instance, if the SNP being analyzed could include a G or an A, the two ASO probes would have the identical sequence to one another except at the region which is complementary to the SNP and at which point the oligonucleotide will include a C or a T, respectively. Thus, an ASO is an oligonucleotide which includes one of two alternative nucleotides at a polymorphic site within its nucleotide sequence. In some embodiments, it is preferred that the oligonucleotide include only a single mismatched nucleotide residue namely the polymorphic residue, relative to an allele of a SNP. In other cases, however, the oligonucleotide may contain additional nucleotide mismatches such as neutral bases or may include nucleotide analogs. In preferred embodiments, the SNP-ASO is composed from about 10 to 50 nucleotide residues. In more preferred embodiments, it is composed of from about 10 to 25 nucleotide residues.

Oligonucleotides may be purchased from commercial sources such as Genosys, Inc., Houston, Texas or, alternatively, may be synthesized de novo on an Applied Biosystems 381A DNA synthesizer or equivalent type of machine.

The ASO probes are labelled. The oligonucleotides may be labeled by any method known in the art. One preferred method is end-labeling, which can be performed as described in Maniatis et al., "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratories, Cold Spring Harbor, New York (1982). The label may be added directly to the ASO during synthesis of the oligonucleotide. For instance, a PCR reaction performed using labeled primers or labeled nucleotides will produce a labeled product. Labeled nucleotides (e.g., fluorescein-labeled CTP) are commercially available. Methods for attaching labels to nucleic acids are well known to those of ordinary skill in the art and, in addition to the PCR method, include, for example, nick translation and end-labeling.

Additionally, the methods may utilize two or more fluorescent dyes which can be spectrally differentiated to reduce the number of samples to be analyzed. For instance, if four fluorescent dyes having spectral distinctions (e.g., ABI Prism dyes 6-FAM, HEX, NED, ROX) are used. Then four hybridization reactions can be carried out under a single hybridization condition.

When the two ASO probes are used to contact a single DNA fragment in a single sample, the two probes have a spectrally distinct label. When separate samples are contacted with the two ASO probes individually, the probes may have different or the same labels. A label may be a light-emissive label, radioactive label, etc. Light-emissive labels can be added to the ASO probe or may be naturally-occurring within the probe. For instance, some bases of dinucleotide are naturally-occurring light-emissive labels. In the case when a naturally-occurring light-emissive label is used, an extrinsic label does not need to be added to the ASO probe. Light-emissive labels, which can be added to ASO probes, include fluorophors and quenchers, light-scattering particles (such as gold particles which scatter light), etc. Radioactive labels include, but are not limited to, ^{3}H, ^{32}P, and ^{35}S. The use of each of these types of labels is well known to those of ordinary skill in the art.

If the step of contacting the DNA fragment with the two ASO probes is performed in solution, either the method must be performed on separate samples or should involve a step or label which can distinguish the two probes and the amount of probes which are hybridized. If the method is performed on a surface, the DNA fragment (or ASO probe) attached to the surface may be attached in two discrete areas, or the ASO probes may have distinct labels. For instance, one of the labelled probes may include a label which is stimulated in the red wave length of light to produce a signal detected in the green wave length. The second labelled probe may include a label which is stimulated in the red wave length to produce a signal detected in the yellow wave length or alternatively may be a label which is stimulated in the orange wave length to emit light in the yellow wave length. When the sample is subjected to light within the red wave length spectrum, it will emit green light and optionally yellow light, depending on the second label. Alternatively, each of the two labelled probes can be labelled with a molecule that is stimulated by distinct wave lengths of light, but each emit in a difference spectrum.

A fluorescent label or fluorophore is a substance which is capable of exhibiting fluorescence within a detectable range. Fluorophores include, but are not limited to, fluorescein, isothiocyanate, fluorescein amine, eosin, rhodamine, dansyl, umbelliferone, 5-carboxyfluorescein (FAM), 2'7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE), rhodamine, 6 carboxyrhodamine (R6G), N,N,N',N'-tetramethyl-6-carboxyrhodamine

(TAMRA), 6-carboxy-X-rhodamine (ROX), 4-(4'-dimethylaminophenylazo) benzoic acid (DABCYL), 5-(2'-aminoethyl) aminonaphthalene-1-sulfonic acid (EDANS), 4-acetamido-4'-isothiocyanatostilbene-2, 2'disulfonic acid, acridine, acridine isothiocyanate, r-amino-N->3-vinylsulfonyl)phenyl!naphthalimide-3,5, disulfonate (Lucifer Yellow VS), N-(4-anilino-1-naphthyl)maleimide, anthranilamide, Brilliant Yellow, coumarin, 7-amino-4-methylcoumarin, 7-amino-4-trifluoromethylcouluarin (Coumaran 151), cyanosine, 4', 6-diaminidino-2-phenylindole (DAPI), 5', 5"-diaminidino-2-phenylindole (DAPI), 5', 5"-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red), 7-diethylamino-3- (4'-isothiocyanatophenyl) -4-methylcoumarin diethylenetriamine pentaacetate, 4,4'-diisothiocyanatodihydro-stilbene-2, 2'-disulfonic acid, 4,4'-diisothiocyanatostilbene-2, 2'-disulfonic acid, 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC), eosin isothiocyanate, erythrosin B, erythrosin isothiocyanate, ethidium, 5-(4,6-dichlorotriazin-2-yl) aminofluorescein (DTAF), QFITC (XRITC), fluorescamine, IR144, IR1446, Malachite Green isothiocyanate, 4-methylumbelliferone, ortho cresolphthalein, nitrotyrosine, pararosaniline, Phenol Red, B-phycoerythrin, o-phthaldialdehyde, pyrene, pyrene butyrate, succinimidyl 1-pyrene butyrate, Reactive Red 4 (Cibacron . RTM. Brilliant Red 3B-A), lissamine rhodamine B sulfonyl chloride, rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101, (Texas Red), tetramethyl rhodamine, tetramethyl rhodamine isothiocyanate (TRITC), riboflavin, rosolic acid, and terbium chelate derivatives.

Fluorescence is measured using a fluorometer. The optical emission from the fluorescence molecule can be detected by the fluorometer and processed as a signal. When fluorescence is being measured in a sample fixed to various portions of the surface, the surface can be moved using a multi-access translation stage in order to position the different areas of the surface, such that the signal can be collected. Many types of flourometers have been developed. For instance, a new sensitive instrument for measuring FRET is described in U.S. Patent No. 5,911,952.

The genotype is analyzed by contacting a DNA fragment with the two ASO probes. The DNA fragment is any nucleic acid sample in which a potential SNP exists.

For instance, the DNA fragment may be an isolated genome, a portion of an isolated genome, or a PCR amplified region of DNA. An isolated genome consists of all of the DNA material from a particular organism, i.e., the entire genome. A portion of an isolated genome, which is referred to herein as a reduced complexity genome (RCG), is a plurality of DNA fragments within an isolated genome but which does not include the entire genome. An amplified region of DNA is a region of DNA which is amplified using polymerase chain reaction (PCR). The PCR amplification is performed by preparing primers that flank the DNA region to be amplified and performing a series of annealing, extension, and separation reactions to produce multiple copies of the DNA region of interest. One advantage of using PCR amplified DNA is that the DNA fragment includes multiple copies of the polymorphic region.

Genomic DNA comprises the entire genetic component of a species excluding, applicable, mitochondrial and chloroplast DNA. Of course, the methods of the invention can also be used to analyze mitochondrial, chloroplast, etc., DNA as well. Depending on the particular species of the subject being analyzed, the genomic DNA can vary in complexity. For instance, species which are relatively low on the evolutionary scale, such as bacteria, can have genomic DNA, which is significantly less complex than species higher on the evolutionary scale. Bacteria, such as *E coli* have approximately 2.4 x 10^9 grams/mol of haploid genome, and bacterial genomes having a size of less than about 5 million base pairs (5 megabases) are known. Genomes of intermediate complexity, such as those of plants, for instance, rice, have a genome size of approximately 700-1000 megabases. Genomes of highest complexity, such as maize or humans, have a genome size of approximately 10^9-10^{11}. Humans have approximately 7.4 x 10^{12} grams/mol of haploid genome.

A "RCG" as used herein is a reproducible fraction of an isolated genome which is composed of a plurality of DNA fragments. The RCG can be composed of random or non-random segments or arbitrary or non-arbitrary segments. The term "reproducible fraction" refers to a portion of the genome which encompasses less than the entire native genome. If a reproducible fraction is produced twice or more using the same experimental conditions the fractions produced in each repetition include at least 50% of the same sequences. In some embodiments the fractions include at least 70%, 80%, 90%,

95%, 97%, or 99% of the same sequences, depending on how the fractions are produced. For instance, if a RCG is produced by PCR another RCG can be generated under identical experimental conditions having at a minimum greater than 90% of the sequences in the first RCG. Other methods for preparing a RCG such as size selection are still considered to be reproducible but often produce less than 99% of the same sequences.

A "plurality" of elements, as used throughout the application refers to 2 or more of the element. A "DNA fragment" is a polynucleotide sequence obtained from a genome at any point along the genome and encompassing any sequence of nucleotides. The DNA fragments of the invention can be generated according to any one of two types mechanisms, and thus there are two types of RCGs, PCR-generated RCGs and native RCGs.

The nucleic acid sample may be prepared using conventional PCR amplification of a polymorphic locus from a genomic DNA sample using known primers. Alternatively PCR-generated RCGs are randomly primed. That is, each of the polynucleotide fragments in the PCR-generated RCG all have common sequences at or near the 5' and 3' end of the fragment (When a tag is used in the primer, all of the 5' and 3' ends are identical. When a tag is not used the 5' and 3' ends have a series of N's followed by the TARGET sequence (reading in a 5' to 3' direction). The TARGET sequence is identical in each primer, with the exception of multiple-primed DOP-PCR) but the remaining nucleotides within the fragments do not have any sequence relation to one another. Thus, each polynucleotide fragment in a RCG includes a common 5' and 3' sequence which is determined by the constant region of the primer used to generate the RCG. For instance, if the RCG is generated using DOP-PCR (described in more detail below) each polynucleotide fragment would have near the 5' or 3' end nucleotides that are determined by the "TARGET nucleotide sequence". The TARGET nucleotide sequence is a sequence which is selected arbitrarily but which is constant within a set or subset (e.g. multiple primed DOP-PCR) of primers. Thus, each polynucleotide fragment can have the same nucleotide sequence near the 5' and 3' end arising from the same TARGET nucleotide sequence. In some cases more than one primer can be used to generate the RCG. When more than one primer is used, each member of the RCG would

have a 5' and 3' end in common with at least one other member of the RCG and, more preferably, each member of the RCG would have a 5' and 3' end in common with at least 5% of the other members of the RCG. For example, if a RCG is prepared using DOP-PCR with 2 different primers having different TARGET nucleotide sequences, a population containing of four sets of PCR products having common ends could be generated. One set of PCR products could be generated having the TARGET nucleotide sequence of the first primer at or near both the 5' and 3' ends and another set could be generated having the TARGET nucleotide sequence of the second primer at or near both the 5' and 3' ends. Another set of PCR products could be generated having the TARGET nucleotide sequence of the second primer at or near the 5' end and the TARGET nucleotide sequence of the first primer at or near the 3' end. A fourth set of PCR products could be generated having the TARGET nucleotide sequence of the second primer at or near the 3' end and the TARGET nucleotide sequence of the first primer at or near the 5' end. The PCR generated genomes are composed of synthetic DNA fragments.

The DNA fragments of the native RCGs have arbitrary sequences. That is, each of the polynucleotide fragments in the native RCG do not have necessarily any sequence relation to another fragment of the same RCG. These sequences are selected based on other properties, such as size or, secondary characteristics. These sequences are referred to as native RCGs because they are prepared from native nucleic acid preparations rather than being synthesized. Thus they are native-non-synthetic DNA fragments. The fragments of the native RCG may share some sequence relation to one another (e.g. if produced by restriction enzymes). In some embodiments they do not share any sequence relation to one another.

In some preferred embodiments, the RCG includes a plurality of DNA fragments ranging in size from approximately 200 to 2,000 nucleotide residues. In a preferred embodiment, a RCG includes from 95 to 0.05% of the intact native genome. The fraction of the isolated genome which is present in the RCG of the invention represents at most 90% of the isolated genome, and in preferred embodiments, contains less than 50%, 40%, 30%, 20%, 10%, 5%, or 1% of the genome. A RCG preferably includes between 0.05

and 1% of the intact native genome. In a preferred embodiment, the RCG encompasses 10% or less of an intact native genome of a complex organism.

Several methods can be used to generate PCR-generated RCG including IRS-PCR, AP-PCR, DOP-PCR, multiple primed PCR, adaptor-PCR and multiple-primed-DOP-PCR. Hybridization conditions for particular PCR methods are selected in the context of the primer type and primer length to produce to yield a set of DNA fragments which is a percentage of the genome, as defined above. PCR methods have been described in many references, see e.g., US Patent Nos. 5,104,792; 5,106,727; 5,043,272; 5,487,985; 5,597,694; 5,731,171; 5,599,674; and 5,789,168. Basic PCR methods have been described in e.g., Saiki et al., Science, 230: 1350 (1985) and U.S. Pat. Nos. 4,683,195, 4,683,202 (both issued Jul. 18, 1987) and U.S. Pat. No. 4,800,159 (issued Jan. 24, 1989).

Another method for generating RCGs is based on the development of native RCGs. Several methods can be used to generate native RCGs, including DNA fragment size selection, isolating a fraction of DNA from a sample which has been denatured and reannealed, pH-separation, separation based on secondary structure, etc.

Size selection can be used to generate a RCG by separating polynucleotides in a genome into different fractions wherein each fraction contains polynucleotides of an approximately equal size. One or more fractions can be selected and used as the RCG. The number of fractions selected will depend on the method used to fragment the genome and to fractionate the pieces of the genome, as well as the total number of fractions. In order to increase the complexity of the RCG, more fractions are selected. One method of generating a RCG involves fragmenting a genome into arbitrarily size pieces and separating the pieces on a gel (or by HPLC or another size fractionation method). A portion of the gel is excised, and DNA fragments contained in the portion are isolated. Typically, restriction enzymes can be used to produce DNA fragments in a reproducible manner.

Different nucleic acid sources may be used to generate RCGs. For instance, mitochondrial DNA can be isolated and used as the source of the RCG.

Separation based on secondary structure can be accomplished in a manner similar to size selection. Different fractions of a genome having secondary structure can be

separated on a gel. One or more fractions are excised from the gel, and DNA fragments are isolated therefrom.

Another method for creating a native RCG involves isolating a fraction of DNA from a sample which has been denatured and reannealed. A genomic DNA sample is denatured, and denatured nucleic acid molecules are allowed to reanneal under selected conditions. Some conditions allow more of the DNA to be reannealed than other conditions. These conditions are well known to those of ordinary skill in the art. Either the reannealed or the remaining denatured fractions can be isolated. It is desirable to select the smaller of these two fractions in order to generate RCG. The reannealing conditions used in the particular reaction determine which fraction is the smaller fraction. Variations of this method can also be used to generate RCGs. For instance, once a portion of the fraction is allowed to reanneal, the double stranded DNA may be removed (e.g., using column chromatography), the remaining DNA can then be allowed to partially reanneal, and the reannealed fraction can be isolated and used. This variation is particularly useful for removing repetitive elements of the DNA, which rapidly reanneal.

The amount of isolated genome used in the method of preparing RCGs will vary, depending on the complexity of the initial isolated genome. Genomes of low complexity, such as bacterial genomes having a size of less than about 5 million base pairs (5 megabases), usually are used in an amount from approximately 10 picograms to about 250 nanograms. A more preferred range is from 30 picograms to about 7.5 nanograms, and even more preferably, about 1 nanogram. Genomes of intermediate complexity, such as plants (for instance, rice, having a genome size of approximately 700-1,000 megabases) can be used in a range of from approximately 0.5 nanograms to 250 nanograms. More preferably, the amount is between 1 nanogram and 50 nanograms. Genomes of highest complexity (such as maize or humans, having a genome size of approximately 3,000 megabases) can be used in an amount from approximately 1 nanogram to 250 nanograms (e.g. for PCR).

In other aspects of the invention, the nucleic acid sample can be an entire or a portion of an RNA genome. RNA genomes differ from DNA genomes in that they are generated from RNA rather than from DNA. An RNA genome can be, for instance, a cDNA preparation made by reverse transcription of RNA obtained from cells of a subject

(e.g. human ovarian carcinoma cells). Thus, an RNA genome can be composed of DNA sequences, as long as the DNA is derived from RNA. RNA samples can also be used directly.

Each of the types of nucleic acid samples set forth herein is described in more detail in co-pending U.S. Patent Application No. 09/404,912, filed on September 24, 1999 (and related PCT publication WO 00/18960), which is hereby incorporated by reference.

Methods of the invention are useful for identifying genome type information and subjects. A subject, as used herein, refers to any type of DNA containing organism and includes, for example, bacteria, virus, fungi, plants, and animals including vertebrates and invertebrates.

EXAMPLES

Example 1: Quantitative Genotype Analysis Using ASO Hybridization and Calculation of Ratio of Intensity of Signal

Materials and Methods:

1. PCR Amplification

Materials:

PCR MasterMix (Abgene) (1.25 units Taq polymerase, 75 mM Tris-HCl pH 8.8, 20 mM (NH4)SO4, 1.5 mM MgCl2, 0.01% Tween20, 0.2 mM each dNTPs)

Primers (Operon): Diluted to 500 uM in 0.1x TE. (Bold letters in the forward primer 4035-1F and reverse primer 4035-1R sequences indicate the T7 and SP6 sequencing tags respectively.)

4035-1F: **TAATACGACTCACTATAGGG**AGTTTGGGACCTGCAAGAAGTAGAT (SEQ ID No.: 1)

4035-1R: **CGATTTAGGTGACACTATAG**CTGTCAAGTATTTCTCCGCAGCATA (SEQ ID No.: 2)

CEPH DNA (Coriell Institute for Medical Research) of 35 people:

Repository No.		
	NA12565	NA11039B
	NA12566	NA11040
	NA12567	NA11041
	NA12568	NA04479

NA12570	NA04477A
NA12571	NA05993
NA12572	NA05995A
NA05542	NA05987A
NA04289	NA05901A
NA05376	NA05918A
NA05536	NA06001A
NA04290	NA05997A
NA04852	NA06003
NA04854	NA06005A
NA04690A	NA05963C
NA11037	NA05961A
NA11038	NA060130
NA11104A	

Methods:

1. In sterile 1.8 mL microfuge tubes, the following components were mixed:

	1 rxn.	Final
MasterMix (~1.1x)	45 uL	1x
(500 uM) Primer 4035-1F	0.1 uL	1 uM
(500 uM) Primer 4035-1R	0.1 uL	1 uM

2. The CEPH DNA was diluted to 10ng/uL in sterile water (Sigma).
3. 2 x 5 uL of each DNA sample (i.e. 2 reactions/person) was dispensed into a 0.2 mL skirted Thermo-Fast 96 well PCR plate (ABgene).
4. 45 uL of the MasterMix/primer mixture was added into the wells of the plate and 45 uL of the mixture was dispensed into a well containing no DNA for a negative control.
5. The plate was centrifuged for 1 minute using the fast cool program (1150 rpm) (Eppendorf 5410R refrigerated centrifuge).
6. The plate was then sealed using Microseal 'A' film (MJ Research).
7. Amplification was carried out using DNA Engine Tetrad Thermalcycler (MJ Research) with the following cycles:

   a. 94°C for 5 min.
   b. 94°C for 1 min.

c. 58°C for 1 min.
d. 72°C for 1.5 min.
e. Repeat steps b-d 39 times.
f. 72°C for 10 min.

8. A 5 uL sample was withdrawn from each reaction and analyze it by gel electrophoresis.

2. PCR Product Purification

1. The 2 reactions were combined for each person (100 uL total).
2. Each reaction mixture was transferred to a MultiScreen-PCR plate (Millipore).
3. The plate was placed on top of a Multiscreen vacuum manifold.
4. A vacuum was applied at 10 inches Hg for 10 minutes until dry.
5. The Multiscreen plate was removed from manifold and add 100 uL of sterile water to each well.
6. The samples were resuspended by shaking the plate vigorously on a rotary shaker (~100 rpm) for 5 minutes.
7. The purified sample was retrieved from each well by pipetting.
8. The amount of PCR product was quantitated by spectrophotometer (OD_{260}).
9. The samples were dried using a vacuum dessicator (Savant) with medium heat.
10. The DNA was resuspended to ~20 ng/uL in 0.1xTE/50% DMSO and stored at –20°C until use.

3. DNA Printing

1. Using a Hydra96 (Robbins Scientific), 10 uL of each sample was aliquoted to 4 wells of a Thermowell 384-well plate (Corning).
2. Using Clonetracker (Ed. 1.4) software (BioDiscovery), a printing protocol was generated for use on microarrayer (ProSys 5510, Cartesian Technologies), that generates replicates of 4 (i.e. 16 spots/person).
3. The samples were printed onto CMT-GAPS slides (Corning) at humidity level of 60%.
4. The slides were dried at room temperature for 1 hr and then baked in an oven (VWR) at 85°C for 2 hours.

4. Post-Processing Slides

1. Using a hotplate, >2 L of water was boiled in a beaker.
2. The position of array was marked on slides using a "diamond" pen.
3. The slides were placed in a metal slide rack and places in boiling water. The slides were then plunged up and down in the water for 2 min. making sure that the temperature did not fall below 95°C.
4. The slides were transferred to 300 mL of 95% EtOH for 1 min.
5. The slides were centrifuged at 800 rpm for 3 min. at room temperature (Eppendorf 5410 Centrifuge).

5. Hybridization of Slides

Materials:

5M Tetramethylammonium Chloride (TMA) (Sigma)

0.5 M EDTA pH 8.0 (Gibco BRL)

10% Sodium Dodecyl Sulfate (SDS) (Gibco BRL)

20x SSC (Gibco BRL)

Dendrimer (Cy3 Labeled 3DNA, Genisphere)

Allele Specific Oligos (ASO) (Operon): Diluted to 500 pmol/uL in 0.1xTE. (The sequences of den4035-1C and den4035-1G contain the Cy3 RT primer "capture sequence" while snp4035-1C and snp4035-1G do not.)

den4035-1C: GGCCGACTCACTGCGCGTCTTCTGTCCCGCCGCCACAATGAA TGACAT (SEQ ID No.: 3)

den4035-1G: GGCCGACTCACTGCGCGTCTTCTGTCCCGCCGCCACAATCAA TGACAT (SEQ ID No.: 4)

snp4035-1C: GCCACAATGAATGACAT (SEQ ID No.: 5)

snp4035-1G: GCCACAATCAATGACAT (SEQ ID No.: 6)

Slides with DNA printed (2)

500 mL 3M TMACB (made fresh daily) (3M TMAC, 10 mM NaPO4 pH 6.8, 1 mM EDTA, 0.1% SDS).

375 uL 4M TMACB

Methods:

Preform dendrimer with ASO

1. ASOs den4035-1C and den4035-1G were diluted to 0.1 pmol/uL via serial dilutions in H_20.
2. In 1.8 mL microfuge tubes, the following components were mixed:

Tube:	1	2
4M TMACB	15 uL	15 uL
(0.1 pmol/uL) den4035-1C	1 uL	-
(0.1 pmol/uL) den4035-1G	-	1 uL
H_2O	0.5 uL	0.5 uL
Dendrimer	2.5 uL	2.5 uL

3. The tubes were wrapped in foil and incubated at 52°C for 4 hr. with shaking on a rotary shaker at 48 rpm.

Prehybridization and Hybridization of Slides

1. 200 mL of 3M TMACB was poured into a wheaton jar.
2. Slides were placed in a glass slide rack and transferred to the wheaton jar.
3. Incubation was performed at 52°C for 1 hr. with shaking on a rotary shaker at 98 rpm.
4. ASOs snp4035-1C and snp4035-1G were diluted to 0.1 pmol/uL via serial dilutions in H_20.
5. ASOs snp4035-1C and snp4035-1G were diluted to 0.1 pmol/uL via serial dilutions in H_20.
6. 1 uL of (0.1 pmol/uL) snp4035-1G was added to tube 1 and 1 uL of (0.1 pmol/uL) 4035-1C to tube 2.
7. 5 uL of water was pipetted into each well of a hybridization chamber (Corning).
8. The slides were removed from the prehybridization solution. The backside and edges of the slides were blotted with a paper towel and placed in the hybridization chamber.
9. The hybridization solution (20 uL total) was pipetted directly onto the array.
10. The array was covered with a glass coverslip (Corning) and the chambers were closed and incubated overnight at 52°C.

Wash Slides

1. One wheaton jar was preheated to 52°C.
2. 800 mL of 3M TMACB was prepared and 200 mL was preheated to 52°C. The rest was preheated to 30°C.
3. 500 mL of 2x SSC was prepared and preheated to 30°C.
4. 400 mL of 0.2x SSC was prepared and preheated to 30°C.
5. The slides were taken out of the hybridization chambers and the coverslips were removed one at a time by gently dunking the slides in a 200 mL beaker containing 3M TMACB.
6. The slides were then placed in a glass slide rack and put into a wheaton jar containing 200 mL 3M TMACB preheated to 30°C and covered with foil. The slides were washed by placing on a rotary shaker at 98 rpm at 30°C for 15 min. The step was repeated once.
7. 3M TMACB (preheated to 52°C) was poured into preheated wheaton jar.
8. The slides were transferred to the wheaton jar and washed with shaking at 98 rpm at 52° C for 15 min.
9. The slides were washed as above in 2x SSC for 2x 5 min. at 30°C.
10. The slides were then washed as above in 0.2x SSC for 5 min. at 30°C.
11. The slides were centrifuged at 400 rpm for 12 min. at room temperature until dry.

Scanning Slides and Analysis

1. The slides were loaded in a cassette of ArrayWorx Microarray Scanner (Applied Precision) equipped with an Arc lamp source and a CCD camera.
2. The slides were scanned with 5 um resolution, 4 sec. exposure, using a CY3/CY3 filter set with excitation/emission wavelengths of 548/595 nm.
3. The genotypes of the people were determined by comparing both slides simultaneously. The color of each slide was set to be different from one another. The slides were overlayed to generate a picture showing the results of both slides. (ex. The presence of signal on one slide (one allele) indicates that the person is

homozygous for that allele while the presence of signal on both slides indicates that the person is a heterozygote.)

Sequence of the polymorphic region (from NCBI website) in the 4035-1 locus

agtttt gggacctgca
agaagtagat acagagaccg gattaagtgt tcaatgattt cactgggga aacgtgtgag
agaaaatggg actagagcca gagaggccca gagtggcctc agactgcaat gtaagtctga
ccttgaaaaa tcaaacagag agagatcaaa cagcagagta cggcagctgg agcccttggt
caaggatgtt ctctgtagtt ggaacagcaa tgcacattac catggtggcc acaat[G/C]
aatgacatgg aagaaatatc taaaagacaa aatgatcaag tttgcagatg acactgggat
gcaatgctta gaatctggtc ttgcaaggtt ttgaatgtgt tctttttcaa agctgaacat
atacaagagc tccttatatg atcccattag cttcttggga tgcctccccc tattctttca
ttgcagttgt ataatcagaa ttatattttg acaggtatga gaaattatgc tgcggagaaa
tacttgacag (SEQ ID No.: 7)

Results:

A microarray chip was spotted with DNA fragments of several subjects. Multiple reactions were performed for each subject. Thus the DNA fragments from each subject were coated onto the chip in 16 discrete spots. The samples were then contacted with the two ASO probes described above on two separate but duplicate surfaces. The intensity of fluorescence emitted from the hybridized probes was determined separately for each ASO probe. The fluorescence emitted was captured as an image and is depicted in figures 1 and 2. The red light emitted in figure 1 corresponds to the fluorescence resulting from hybridized ASO probe #1. The green light emitted in figure 1 corresponds to the fluorescence resulting from hybridized ASO probe #1.

Figure 3 is a computer overlay image of the images of the microarrays from Figure 1 and Figure 2. From an initial review of Figure 3, a preliminary assessment of which samples are homozygous and which are heterozygous can be made. In general the samples which appear red or green represent homozygous DNA samples and the samples that appear to be a combination of red and green are heterozygous. An analysis of the intensity of the two signals and a comparison of that intensity a more accurate, quantitative value of the genotype of each sample.

Figure 4 is a computer generated image depicting the ratio of fluorescent label detected from the first and second ASO probes, following the overlay shown in Figure 3. The regions of the image consisting of entirely green circles correspond to a DNA sample

which is homozygous for one allele. Those regions of the image consisting of entirely red circles correspond to a DNA sample which is homozygous for the other allele. The mixed color circles correspond to a DNA sample which is heterozygous. The results of the intensity analysis depicted in figure 4 can be presented in the form of a histogram. Thus, Figure 5 is a histogram depicting the quantitative values of the ratio of the fluorescent labels from the two ASO probes. The values ranging from 10^{-1} to 10^{1} represent the DNA fractions which are heterozygous. The values ranging from 10^{-4} to 10^{-1} represent the DNA fractions which are homozygous for one allele. The values ranging from 10^{1} to 10^{3} represent the DNA fractions which are homozygous for the other allele.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not limited in scope by the examples provided, since the examples are intended as illustrations of various aspects of the invention and other functionally equivalent embodiment are within the scope of the invention. Various modifications of the invention in addition to those shown in and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the inventive claims. The advantages and objectives of the invention are not necessarily encompassed by each embodiment of the invention. All references, patents, and patent publications which are cited in this application are incorporated in their entirety herein by reference.

I claim:

CLAIMS

1. A method for analyzing a genotype, comprising:

contacting a DNA fragment with two ASO probes, each ASO probe representing one allele of a SNP, and each ASO probe having a different label, and

measuring the ratio of the ASO probes specifically hybridized to the DNA fragment to analyze the genotype.

2. The method of claim 1, wherein an approximately equivalent amount of the two ASO probes are used.

3. The method of claim 1, wherein the DNA fragment is fixed to a surface.

4. The method of claim 3, wherein the surface is a multiwell dish.

5. The method of claim 3, wherein the surface is a chip.

6. The method of claim 3, wherein the surface is a slide.

7. The method of claim 3, wherein the surface is a bead.

8. The method of claim 1, wherein each of the ASO probes is hybridized to the DNA fragment simultaneously with one another.

9. The method of claim 1, wherein the ASO probes are hybridized to the DNA fragment consecutively.

10. The method of claim 1, wherein at least one of the ASO probes contains a fluorescent label or quencher.

11. The method of claim 10, wherein both of the ASO probes contain a fluorescent label or quencher, the fluorescent label or quencher of the two ASO probes, being distinct from one another.

12. The method of claim 2, wherein the DNA fragment is prepared by PCR amplification of a polymorphic locus from a genomic DNA sample.

13. The method of claim 1, wherein the DNA fragment is a reduced complexity genome.

14. The method of claim 1, wherein the DNA fragment is genomic DNA.

15. The method of claim 1, wherein the DNA fragment is a mitochondrial genome.

16. A method for analyzing a genotype, comprising:

contacting a first DNA fragment with a first ASO probe representing one allele of a SNP,

contacting a second DNA fragment with a second ASO probe representing a second allele of the SNP, and

measuring the ratio of the ASO probes specifically hybridized to the DNA fragment to analyze the genotype.

17. The method of claim 16, wherein an approximately equivalent amount of the two ASO probes are used.

18. The method of claim 16, wherein the first and second DNA fragments are fixed to a surface.

19. The method of claim 18, wherein the surface is a multiwell dish.

20. The method of claim 18, wherein the surface is a chip.

21. The method of claim 18, wherein the surface is a slide.

22. The method of claim 18, wherein the surface is a bead.

23. The method of claim 20, wherein the first and second DNA fragments are fixed on discrete locations of the chip and each of the ASO probes is hybridized to the DNA fragments simultaneously with one another.

24. The method of claim 20, wherein the first and second DNA fragments are fixed on discrete locations of the chip and the ASO probes are hybridized to the DNA fragments consecutively.

25. The method of claim 16, wherein at least one of the ASO probes contains a fluorescent label or quencher.

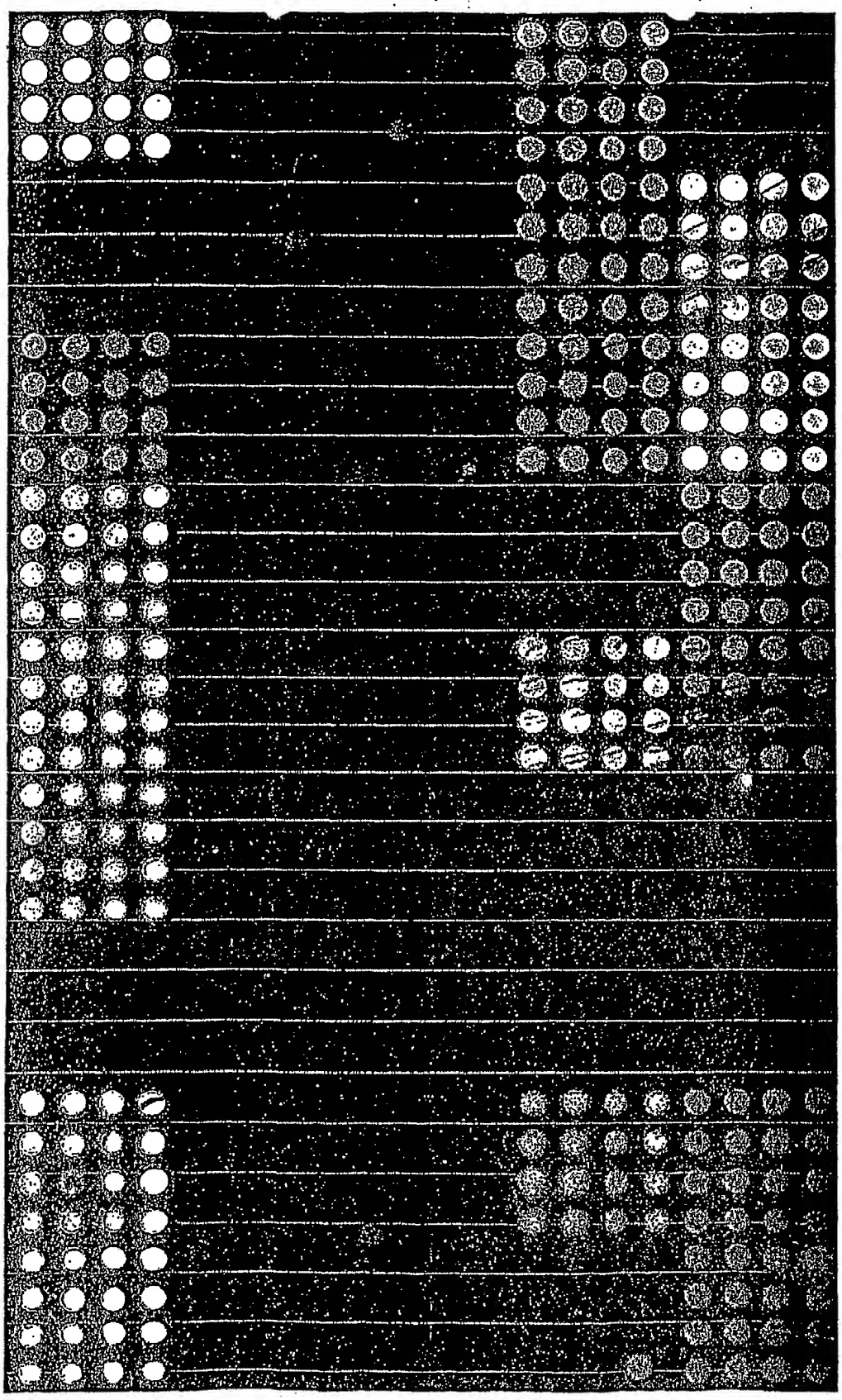
26. The method of claim 25, wherein both of the ASO probes contain a fluorescent label or quencher, the fluorescent label or quencher of the two ASO probes, being distinct from one another.

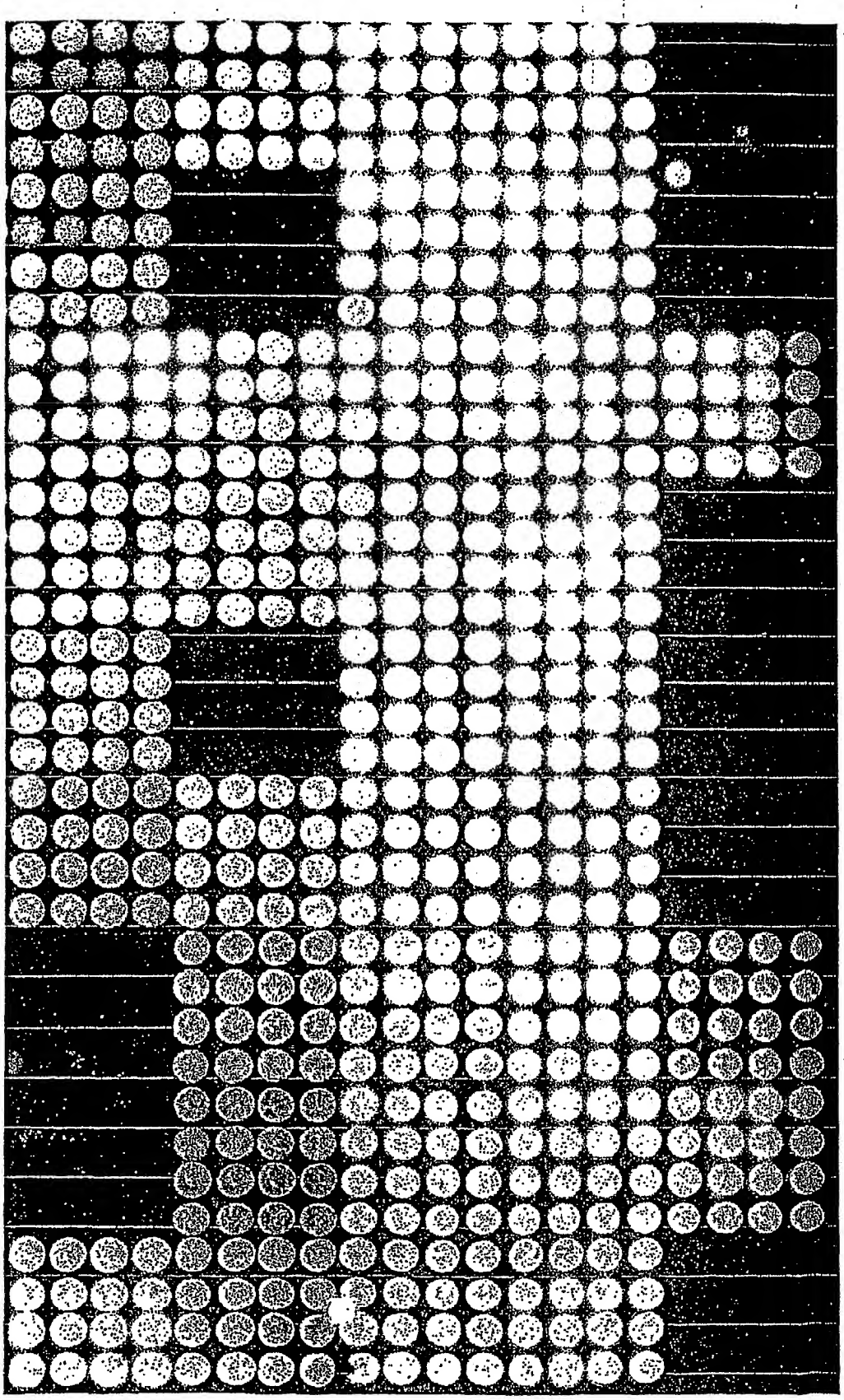
27. The method of claim 17, wherein the DNA fragments are prepared by PCR amplification of a polymorphic locus from a genomic DNA sample.

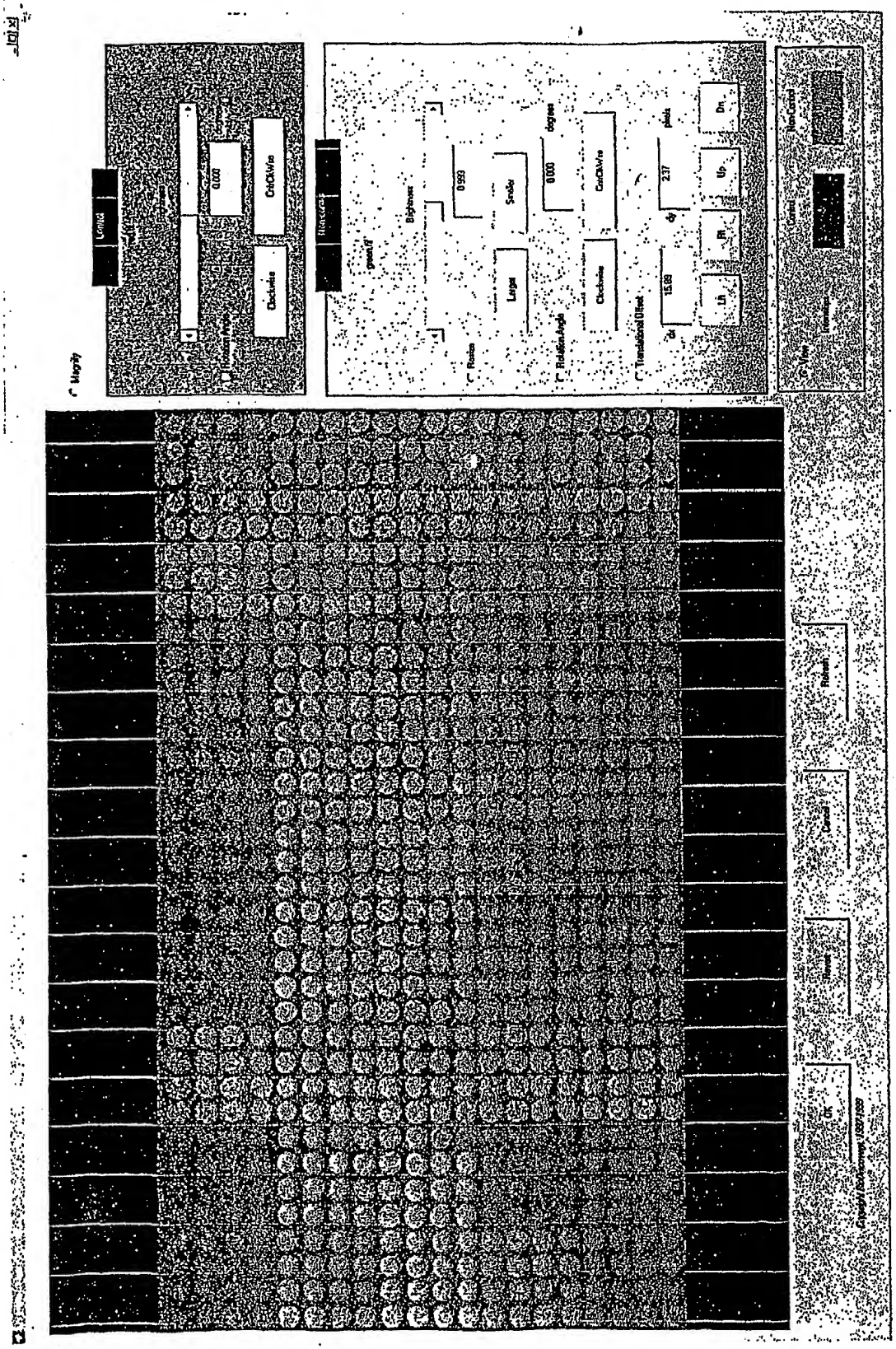
28. The method of claim 16, wherein the DNA fragments are a reduced complexity genome.

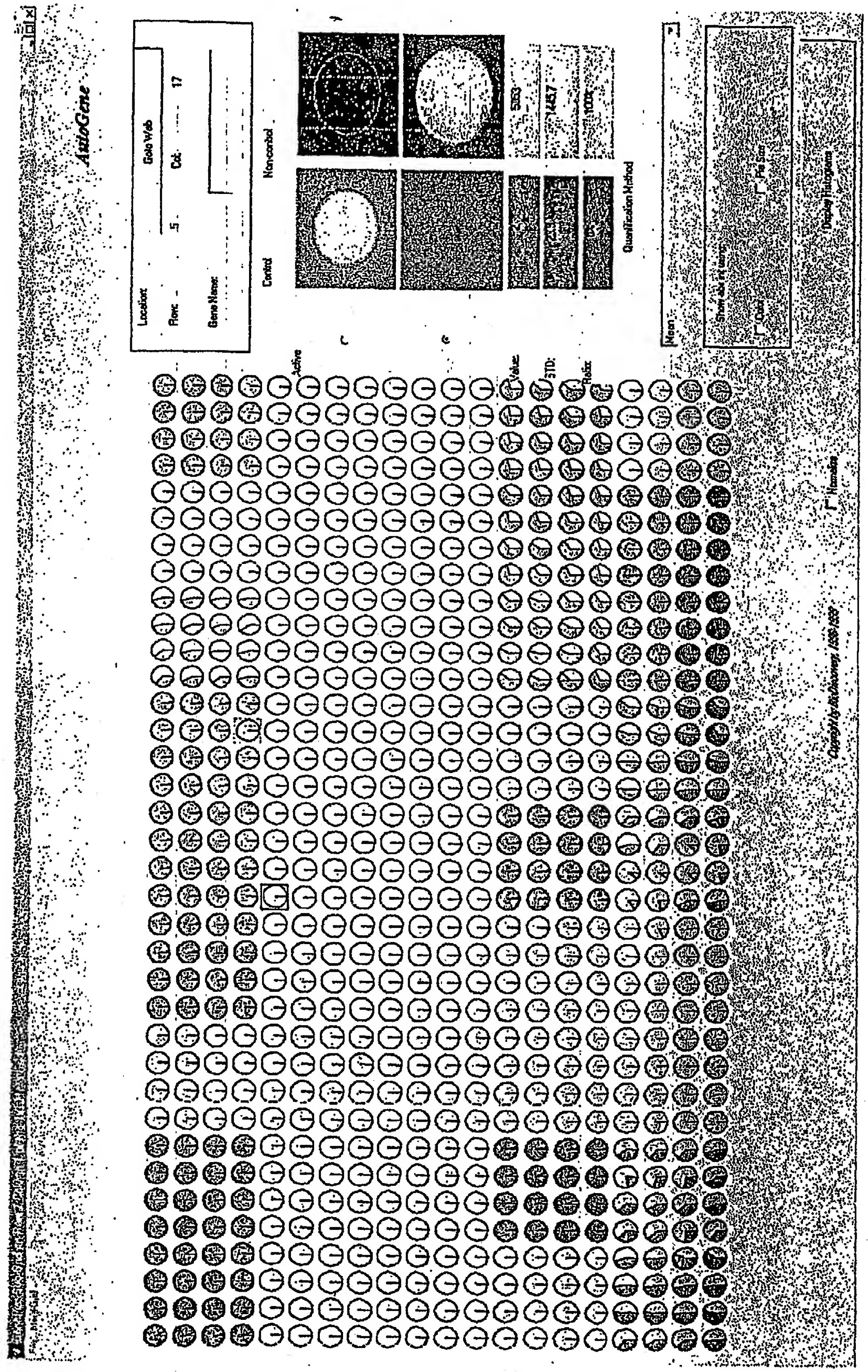
29. The method of claim 16, wherein the DNA fragments are genomic DNA.

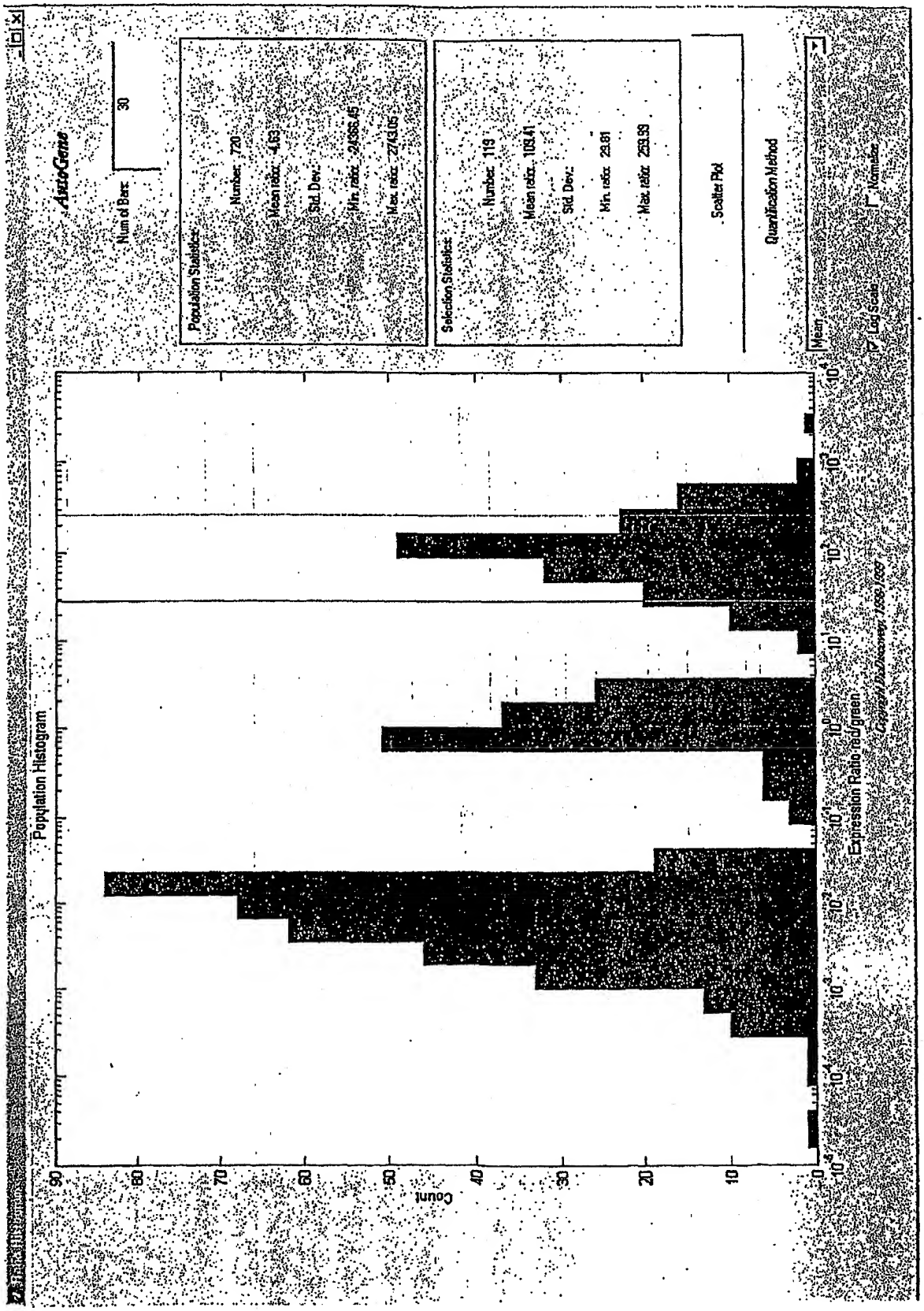
30. The method of claim 16, wherein the DNA fragments are a mitochondrial genome.






Control
0.000
CntrClkWise
Clockwise
Horizontal Mode
Magnify
Horizontal
Brightness
0.900
Radius
Small
Large
Rotation Angle
0.000
degrees
CntrClkWise
Clockwise
Translational Offset
15.00
dx
2.37
pixels
dy
Lft
Rt
Up
Dn
Non-Control
View
Information
Reset
Cancel
OK


AutoGene
Location:
Gene Web
Row:
5
Col:
17
Gene Name:
Control
Non-control
Quantification Method
Active
Value:
STD:
Ratio:
Mean
Display Histogram


AutoGene
Num of Bars
30
Population Statistics
Number: 720
Mean ratio: 4.63
Std. Dev.:
Min. ratio: 24366.46
Max. ratio: 2743.06
Selection Statistics
Number: 119
Mean ratio: 109.41
Std. Dev.:
Min. ratio: 29.81
Max. ratio: 259.99
Scatter Plot
Quantification Method
Mean
Log Scale
Normalize
Population Histogram
Count
Expression Ratio: red/green

SEQUENCE LISTING

<110> PolyGenyx, Inc

<120> Methods and Products Related to High Throughput Genotype Analysis

<130> P0715/7004PCT(HCL/MXA)

<150> US 60/231,281

<151> 2000-09-08

<160> 7

<170> PatentIn version 3.0

<210> 1

<211> 45

<212> DNA

<213> Artificial Sequence

<400> 1
taatacgact cactataggg agtttgggac ctgcaagaag tagat 45

<210> 2

<211> 45

<212> DNA

<213> Artificial Sequence

<400> 2
cgatttaggt gacactatag ctgtcaagta tttctccgca gcata 45

<210> 3

<211> 48

<212> DNA

<213> Artificial Sequence

<400> 3
ggccgactca ctgcgcgtct tctgtcccgc cgccacaatg aatgacat 48

<210> 4

<211> 48

<212> DNA

<213> Artificial Sequence

<400> 4
ggccgactca ctgcgcgtct tctgtcccgc cgccacaatc aatgacat 48

<210> 5

<211> 17

<212> DNA

<213> Homo sapiens

<400> 5
gccacaatga atgacat 17

<210> 6

<211> 17

<212> DNA

<213> Homo sapiens

<400> 6
gccacaatca atgacat 17

<210> 7

<211> 501

<212> DNA

<213> Homo sapiens

<400> 7
agtttgggac ctgcaagaag tagatacaga gaccggatta agtgttcaat gatttcactg 60

ggggaaacgt gtgagagaaa atgggactag agccagagag gcccagagtg gcctcagact 120

gcaatgtaag tctgaccttg aaaaatcaaa cagagagaga tcaaacagca gagtacggca 180

gctggagccc ttggtcaagg atgttctctg tagttggaac agcaatgcac attaccatgg 240

tggccacaat saatgacatg gaagaaatat ctaaaagaca aaatgatcaa gtttgcagat 300

gacactggga tgcaatgctt agaatctggt cttgcaaggt tttgaatgtg ttctttttca 360

aagctgaaca tatacaagag ctccttatat gatcccatta gcttcttggg atgcctcccc 420

ctattctttc attgcagttg tataatcaga attatatttt gacaggtatg agaaattatg 480

ctgcggagaa atacttgaca g 501

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/27655

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68
US CL : 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 91.1, 287.2; 536/23.1, 23.3, 23,33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST (US PAT, US PreGt; EPO, JPO, Derwent)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US A 5,834,181 A (SHUBER) 10 November 1998, see columns 5-7 and 9.	1-30
Y	WO 99/01576 A1 (UNIVERSITY OF BRISTOL) 14 January 1999; see pages 5, 8, 12, and 14.	1-30
Y	US 5,633,134 A (SHUBER) 27 May 1997, see column 3.	1-30

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 November 2001 (02.11.2001)	20 DEC 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Bradley L. Sisson Telephone No. 703-308-1234

Form PCT/ISA/210 (second sheet) (July 1998)